Amendments to the Specification:

Please replace the paragraph of page 8, line 17 with the following amended paragraph:

FIGURE 3 shows an embodiment of a multi-tier internet database system that is useful for some embodiments of the invention (For a description of an Internet database platform, *see, e.g.,* the Java™ 2 Platform, Enterprise Edition Application Programming Model described by Sun Microsystems, see java.sun.com/j2ee/apm/, ~~http://java.sun.com/j2ee/apm/,~~ last accessed on December 14, 2000). The database (301), *e.g,* a gene expression database or a genotyping database, and system external to the data (302) reside in one or several data servers which constitute the data server tier.

Please replace the paragraph of page 9, line 1 with the following amended paragraph:

Java enabled application servers (303) contain distributed, reusable business components housed in either a Java Common Object Request Broker Architecture (CORBA) Object Request Broker (ORB) or an Enterprise JavaBean (EJB) server. For a description of the distribute object technology, see, *e.g.*, specifications and other documents at the web-site of the Object Management Group (OMG), ~~http://www.omg.org,~~ all incorporated herein by reference for all purposes.

Please replace the paragraph of page 9, line 7 with the following amended paragraph:

The business components publish their data and services to Graphic User Interface (GUI) clients or other servers via component application programming interfaces (APIs) like CORBA and EJB, messaging APIs like Java Messenger Service (JMS), or data exchange formats like Extensible Markup Language (XML). The April 2000 specification of the XML is available at the World Wide Web consortium website ~~http://www.w3.org~~ and is incorporated herein by reference for all purposes.

Please replace the paragraph of page 9, line 17 with the following amended paragraph:

Thin client HTML interfaces (305) are dynamically generated by Java enabled web servers (304) using, for example, JavaServer Pages (JSP) and Java Servlet standards ~~(www.javasoft.com)~~. More functionally rich and productive thick clients are assembled from libraries of reusable JavaBeans. The Java clients can run either as applets augmenting HTML within a Java enabled browser (306) or as applications running independently on the desktop (307). Java clients typically connect to application servers via Internet Inter-ORB Protocol (IIOP) or directly to data servers using JDBC.

Please replace the paragraph on page 10, line 4 with the following amended paragraph:

Relational databases store all of their information in groups known as tables. Each database can contain one or more of these tables. A relational database management

system (RDBMS) can also manage many individual underlying databases, with each one of these databases containing many tables. These tables are related to each other using some type of common element. A table can be thought of as containing a number of rows and columns. Each individual element stored in the table is known as a column. Each set of data within the table is known as a row. There are a number of commercial or public domain relational DBMS (RDBMS) such as Oracle ~~(www.oracle.com)~~, Sybase ~~(www.sybase.com)~~, Microsoft® SQL server and MySQL ~~(www.mysql.com)~~.

Please replace the paragraph on page 10, line 13 with the following amended paragraph:

One preferred language for managing relational database is the SQL. Structured Query Language (SQL) is an American National Standard Institute (ANSI) standard computer programming language. SQL is useful for querying and managing relational databases. The ANSI standard for SQL (SQL-92, available at the ANSI website, ~~www.ansi.org~~, last visited on December 14, 2000 and is incorporated herein by reference for all purposes) specifies a core syntax for the language itself. For a detailed description of the SQL language, see, *e.g.*, The Practical SQL Handbook : Using Structured Query Language by Judith S. Bowman, *et al.*, Addison-Wesley Pub Co; ISBN: 0201447878, which is incorporated herein by reference for all purposes. Many embodiments of the invention employ SQL for query and database management.

Please replace the paragraph on page 15, line 8 with the following amended paragraph:

Typically, a nucleic acid sample is labeled with a signal moiety, such as a fluorescent label. The sample is hybridized with the array under appropriate conditions. The arrays are washed or otherwise processed to remove non-hybridized sample nucleic acids. The hybridization is then evaluated by detecting the distribution of the label on the chip. The distribution of label may be detected by scanning the arrays to determine fluorescence intensity distribution. Typically, the hybridization of each probe is reflected by several pixel intensities. The raw intensity data may be stored in a gray scale pixel intensity file. The GATC™ Consortium has specified several file formats for storing array intensity data. The final software specification is available at the website of the GATCC Consortium ~~www.gateconsortium.org~~ and is incorporated herein by reference in its entirety. The pixel intensity files are usually large. For example, a GATC™ compatible image file may be approximately 50 Mb if there are about 5000 pixels on each of the horizontal and vertical axes and if a two byte integer is used for every pixel intensity. The pixels may be grouped into cells (see, GATC™ software specification). The probes in a cell are designed to have the same sequence (i.e., each cell is a probe area). A CEL file contains the statistics of a cell, e.g., the 75th percentile and standard deviation of intensities of pixels in a cell. The 50, 60, 70, 75 or 80th percentile of pixel intensity of a cell is often used as the intensity of the cell.

Please replace the paragraph on page 17, line 3 with the following amended paragraph:

In one aspect of the ~~invneiton~~ invention, a relational data model is designed for the integration of biological knowledge with expression data. Biological knowledge is integrated following the central dogma of biological macromolecules: DNA, mRNA and protein. Database entities were designed to mimic the biological entities, the relationship among entities mimics the relationship among biological macromolecules, for instance, one gene can have many orthologous loci, one locus can produces many transcripts, and one transcript can generate one or more proteins. This data model is also faithful to the way biological knowledge is organized. For example, a protein domain is linked to protein entity because it's a property of protein, gene ontology is associated with the locus entity because it's knowledge developed against a DNA locus.

Please replace the paragraph on page 18, line 12 with the following amended paragraph:

In one aspect of the invention, methods for analyzing gene expression are provided. In some embodiments, the methods include the steps of obtaining expression levels of a plurality of genes; selecting at least one biological characteristic from a plurality of biological characteristics stored in a database; where the biological characteristics comprise genomic information about the genes, structural information about the products of the genes; and biological function of the genes; and analyzing the expression levels according to the selected at least one biological characteristic. The expression levels can be relative or absolute levels of any measurements that can indicate the expression of

genes. For example, the expression levels can be RNA transcript concentrations (micromolar or other units) in a sample; RNA transcript concentrations relative to a particular transcript; protein concentrations in sample etc. One of skill in the art would appreciate that the invention is not limited to any particular measurement of gene expresion or any particular technology for measuring gene expression. However, many embodiments of the invention are particularly suitable for analyzing the expression of a large number of, at least 50, 100, 500, 1000, 5000 and 10,000 genes. The term "biological characteristic, " as used herein, refers broadly to any characteristics that has biological relevancy. For example, a biological characteristic may be chromosomal location, cellular location (particularly for intermediate or final products of gene expression), molecular or cellular functions, structural information (including sequence information, three dimensional structure, protein domains, etc.). In one embodiments, the biological characterstics are described using gene ontology system. The Gene Ontology Consortium (GO) provides a set of standardized vocabulary to describe various biological characteristics. The three organizing principles of GO are molecular function, biological process and cellular component. The current gene ontology information is available at the Gene Ontology Consortum web site ~~at (www.geneontology.com)~~.